

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Publications, Agencies and Staff of the U.S.
Department of Commerce

U.S. Department of Commerce

2012

Analysis of inconsistencies in multi-year gridded quantitative precipitation estimate over complex terrain and its impact on hydrologic modeling

Naoki Mizukami
NOAA, mizukami@ucar.edu

Michael B. Smith
NOAA

Follow this and additional works at: <https://digitalcommons.unl.edu/usdeptcommercepub>



Part of the [Environmental Sciences Commons](#)

Mizukami, Naoki and Smith, Michael B., "Analysis of inconsistencies in multi-year gridded quantitative precipitation estimate over complex terrain and its impact on hydrologic modeling" (2012). *Publications, Agencies and Staff of the U.S. Department of Commerce*. 349.
<https://digitalcommons.unl.edu/usdeptcommercepub/349>

This Article is brought to you for free and open access by the U.S. Department of Commerce at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Publications, Agencies and Staff of the U.S. Department of Commerce by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.



Analysis of inconsistencies in multi-year gridded quantitative precipitation estimate over complex terrain and its impact on hydrologic modeling

Naoki Mizukami ^{a,b,c,*}, Michael B. Smith ^a

^a NOAA, National Weather Service, Office of Hydrologic Development, United States

^b Len Technologies, United States

^c Now at Research Application Laboratory, National Center for Atmospheric Research, United States

ARTICLE INFO

Article history:

Received 22 September 2011

Received in revised form 1 December 2011

Accepted 22 January 2012

Available online 1 February 2012

This manuscript was handled by Konstantine P. Georgakakos, Editor-in-Chief, with the assistance of Emmanouil N. Anagnostou, Associate Editor

Keywords:

Hydrologic modeling

Model calibration

Quantitative precipitation estimate

Time series homogeneity

SUMMARY

It is common that bias trends of long term precipitation data change over time due to various factors such as gauge relocation and changes in data processing methods. Temporal consistency of this error characteristic of precipitation data is as important as accuracy itself for reliable streamflow prediction with a hydrologic model. The main goal of this paper is to illustrate the detection and adverse effect of inconsistent precipitation data on distributed hydrologic model simulations over a mountainous basin in the Sierra Nevada Mountains of California. We used 1-h 4 km gridded precipitation time series intended for the second phase of the Distributed Model Intercomparison Project (DMIP-2), sponsored by the US National Weather Service (NWS). We present various analyses to investigate the consistency of an hourly gridded precipitation time series from October 1988 through September 2006. First, hourly gridded precipitation data were aggregated into monthly mean areal precipitation totals over the basin and compared with basin average totals derived from the Parameter-Elevation Regressions on Independent Slopes Model (PRISM) monthly values. Second, double mass analysis was performed on several discrete locations within the basin using both the DMIP-2 gridded precipitation and PRISM data. In addition, we performed a statistical consistency test on the DMIP-2 time series. Both analyses lead to the conclusion that over the entire analysis time period a clear change in bias in the DMIP-2 data occurred in the beginning of 2003. Moreover, it was found that the PRISM data also exhibit some inconsistency. The inconsistency of two elevation zone mean area precipitation (MAP) time series computed from the DMIP-2 gridded precipitation fields was corrected by adjusting hourly values based on the result from double mass analysis. Model simulations using the adjusted MAP data are improved compared to simulations with the inconsistent MAP input data.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Since any hydrologic model is an imperfect representation of the real world to varying degrees, calibration is an important part of modeling for reliable streamflow simulation and forecasting. During calibration, non-observable model parameters are automatically or manually adjusted using observed input (e.g. precipitation, air temperature) and output (streamflow) data in such a way as to achieve an optimal statistical and visual agreement between simulated and observed streamflow. Acquisition and quality control of historical time series of observed data are the first steps of the calibration process (e.g., Anderson, 2002; Smith et al., 2003). Successful model calibration depends largely on the length and

quality of the observation dataset. In the past, several studies have investigated the impact of the length of calibration period on model calibration and simulation results for a particular basin assuming that input forcing variables were error-free throughout the calibration and validation periods (Brath et al., 2004; Gan et al., 1997; Sorooshian et al., 1983; Yapo et al., 1996). Although those studies used models of various complexities (number of model parameters), types (lumped or distributed), and calibration processes (algorithms for automated calibration and objective function), the general consensus is that the length of calibration datasets should be long enough to encompass average, wet, and dry periods so as to expose the model to a wide range of hydrologic conditions. Thus the desirable calibration period length depends largely on the hydroclimatic conditions particular to the basin.

Among model input data, precipitation is arguably one of the most influential forcings for hydrologic simulation. At the same time, spatial fields of precipitation are perhaps the most challenging dataset to develop. This is particularly true in mountainous re-

* Corresponding author. Now at National Center for Atmospheric Research, Research Application Laboratory, P.O. Box 3000, Boulder, CO 80307, United States. Tel.: +1 303 497 2739.

E-mail address: mizukami@ucar.edu (N. Mizukami).

gions where spatial patterns of precipitation can be extremely heterogeneous. Sparse rain gauge networks make it nearly impossible to retrieve the true spatial distribution of precipitation. Bales et al. (2006) stated that strategic hydrologic instrumentation coupled with remote sensing and ground measurements, in addition to advanced modeling, is needed for complex terrain to retrieve hydro-meteorological data useful for hydrologic predictions. At present, radar products from the Weather Surveillance Radar-1988 Doppler (WSR-88D) radars provide spatial patterns of precipitation over most parts of the nation (Crum and Alberty, 1993). These data have been largely responsible for a resurgence in the development and implementation of distributed models. However, quantitative precipitation estimates (QPE) from radar contain errors, and the reduction of these errors relies on gauge measurements (Fulton et al., 1998; Steiner et al., 1999). With this in mind, radar-based estimates may not be reliable for complex terrain because of sparse gauge networks. In addition, radar products are susceptible to obstruction of radar signals by rugged topography and trees, thus degrading their range and accuracy (Warner et al., 2000; Westrick et al., 1999). Snowfall rate estimates from radar signals have additional uncertainty due to the wide range of snowflake shapes and fall velocities (Rasmussen et al., 2003).

Acknowledging such issues, on-going efforts toward improvement of radar QPE over complex terrain are being made through the National Oceanic and Atmospheric Administration (NOAA) Hydrometeorological Testbed (HMT) experiment around the North Fork of the American River Basin in California (e.g., Gourley et al., 2009; Matrosov, 2010; Matrosov et al., 2009). Despite such significant efforts, hydrologic modeling applications of radar products in mountainous areas may be hampered by insufficient dataset length for model calibration and validation processes, as well as changes in processing algorithms (Reed et al., 2004; Xie et al., 2006; Young et al., 2000). In light of the limitations of current radar-based precipitation estimates, especially in mountainous regions, perhaps the only reliable source for multi-year QPE datasets for hydrologic modeling is from gauge-based historical precipitation data.

When considering the value of long term precipitation datasets for model calibration and validation, whether they are point or gridded data, the key measures of data quality are accuracy and consistency. Gauge measurements of precipitation are often thought of as “true values”; however, they always include measurement errors. In general, gauge measurements are likely to collectively underestimate the true value. If the magnitude of the measurement error is the same at every measurement, this dataset is consistent despite the presence of systematic errors in the dataset. There are a variety of well known systematic errors of gauge measurements which can contribute to the undercatch of precipitation including, but not limited to, the wind field disturbance around the gauge orifice, wetting losses due to water adhering to the gauge surface, evaporation of the water captured in the gauge occurring between the storm events and the timing of measurements, and the treatment of trace precipitation as zero (e.g., Adam and Lettenmaier, 2003; Legates, 1987; Sevruk, 1982).

Underestimation is more severe in snowfall (Pan et al., 2003; Sheffield et al., 2003). This large systematic error in gauge measurements of snowfall greatly hampers the simulation of streamflow and the water balance (Lohmann et al., 2004). Although several studies quantified an adjustment factor to correct or reduce precipitation underestimates for a particular gauge type (Metcalfe et al., 1997; Yang et al., 1998b), estimating the adjustment factor at every gauge requires a wide range of climate and historical gauge information for each gauge, which may not be available. The characteristics of such systematic and random errors can be altered over time due to gauge relocations, observation time changes, instrumentation changes, and changes in the surrounding land-

scape (Chang and Lee, 1974; Yang et al., 1998a). This results in inconsistent time series data. In particular, data inconsistencies caused by gauge relocations become severe in mountainous regions due to the higher spatial variability of meteorological variables there as compared to flat regions (Hamlet and Lettenmaier, 2005).

Recognizing that true precipitation values may never be known, the consistency of input data (i.e. consistent overestimation or underestimation relative to true values throughout the calibration and validation period) is as important as the accuracy of the data to model calibration and subsequent use of the calibrated model (Sutcliffe and Lazenby, 1990). If precipitation data are inaccurate but the measurement error is consistent, the model calibration somewhat compensates for the systematic error by adjusting parameter values to produce reasonable simulations (Andréassian et al., 2001; Brath et al., 2004; Dawdy and Bergmann, 1969; Troutman, 1982, 1983; Xu and Vandewiele, 1994). However, it has been noted calibration in the presence of severe errors in calibration data may result in parameter estimates that are biased or even physically unrealistic (Sorooshian and Gupta, 1983; Troutman, 1985). A model with biased parameters is likely to produce unreliable streamflow forecasts if the error in the input data used for subsequent simulations is not consistent with the calibration data or when the data falls outside of the range used for calibrations. On the other hand, calibrating hydrologic models with inconsistent calibration datasets within which the systematic error changes is difficult (Andréassian et al., 2001).

The primary goal of this paper is to illustrate the significance of consistency in precipitation data for hydrologic modeling. While much research has investigated the impact of data quality on the efficacy of parameter calibration, we address this issue within the confines of analyzing historical gridded precipitation estimates used for distributed model calibration and simulation in regions of complex terrain. The streamflow simulation results presented here are from two mountainous basins used in the second phase of the Distributed Modeling Intercomparison Project (DMIP-2). DMIP-2 (and its predecessor DMIP-1) was initiated to guide the NWS in the use of distributed models for operational river and flash flood forecasting. As such, DMIP-1 and -2 are venues to evaluate the capability of a wide range of existing distributed hydrologic models forced with gridded precipitation data. Interested readers are referred to Smith et al. (2012a,b, 2004) and Reed et al. (2004) for more information on the DMIP series of experiments.

The error consistency of gridded precipitation time series has received much less attention than the analysis of point gage data. To the best of our knowledge, the first to examine gridded precipitation data were Guentchev et al. (2010), who performed several statistical tests to examine the homogeneity of annual gridded precipitation data over the Colorado River Basin. We explored similar approaches to analyze the temporal consistency of hourly gridded data on a monthly basis over a mountainous basin in the Sierra Nevada, California. The results implied a clear link between the timing of changes in streamflow simulation error pattern and precipitation data inconsistencies found in the analysis. Finally, we employed a simple method to correct the inconsistencies found in DMIP-2 precipitation data, and the error in the simulation with the corrected precipitation was examined using a lumped model. The results clearly illustrate the impact of consistency of precipitation data on the simulation results.

The remainder of this paper is organized as follows. Section 2 describes the background of DMIP, which motivated us to initiate this study. Section 3 gives an overview of the study basins and gridded precipitation analyzed in this study. Methods used to detect inconsistencies in the gridded QPE and the results are given in Section 4. In Section 5, a simple method is used to correct the data inconsistency to illustrate how consistency of the input pre-

precipitation data affects hydrologic simulations. Finally a summary of this study and conclusions are provided in Section 6.

2. Background

Prompted by increasing end-user requests for new and finer space-time scale water resources products, and supported by the increasing availability of gridded datasets such as physiographic and climate data, the NWS Office of Hydrologic Development (OHD) is continuing to develop and deploy distributed hydrologic models for operational use. As part of its research and development effort, the NWS OHD organized the DMIP series of experiments to compare a wide range of distributed models developed at academic and research institutions around the world. The project has sought to test the hypothesis that distributed modeling using higher resolution data will produce more accurate streamflow simulations at basin outlets than a lumped hydrologic model (Smith et al., 2004). We have completed the first phase of DMIP (DMIP-1) that evaluated non-mountainous basins (Oklahoma and Arkansas, see special issue in the Journal of Hydrology, Vol. 298, 2004). After DMIP-1, the NWS deployed an operational version of its research distributed hydrologic model in February of 2007 for river stage forecasting in non-snow portions of the United States. A description of the basic components of the OHD distributed model (OHD-DM) is given by Koren et al. (2004).

In an attempt to expand the applicability of distributed modeling into more topographically complex areas of the United States, the scope of DMIP-1 was extended into a second phase (DMIP-2) covering mountainous western areas where hydrologic simulation is more challenging due to the complexity of topography, land cover, and other meteorological variables. During the planning of DMIP-2, there were no established methods to create hourly gridded QPE datasets for mountainous-area distributed hydrologic modeling outside of radar products, which are unreliable in mountainous and snowy regions. To support DMIP-2, hourly gridded precipitation data (hereafter DMIP-2 QPE) were generated from Water Year (WY)-1989 (WY is defined as a period from October 1st of the previous year through September 30th of a specified year) through WY-2005 using available gauge measurements in and around the DMIP-2 study basins (Moreda et al., 2006; Schaake et al., 2004).

In the initial phases of DMIP-2, we tested the OHD-DM over two Sierra Nevada river basins – the North Fork of the American River (NFAR) and the East Fork of the Carson River Basin (EFCR) (Fig. 1; Smith et al., 2009). The results revealed large errors in the streamflow simulation over the study period for the NFAR while the simulation for the EFCR produced a consistent trend in simulation bias (Fig. 2). The simulations for NFAR are consistently underestimated until the end of WY-2002. The calibration period used in this simulation was from October 1988 through September 1997, the rest of the period was used as a validation period. A large underestimation seen in the beginning of 1995 is possibly caused by a recording malfunction at one of the nearby gauges where large rainfall amounts were erroneously recorded as zero during January 1995 (Peter Fickenscher, California-Nevada River Forecast Center – CNRFC, personal communication). Of more significance to hydrologic simulation is the abrupt change in the bias pattern from underestimation to overestimation around the beginning of 2003 for NFAR. A similar, though not as large, cumulative error trend was observed for some Oklahoma basins in DMIP-1. Reed et al. (2004) proposed that this trend change was related to upgrades in the algorithms used to generate radar-based QPE. Based on the DMIP-1 experience, we also suspected that the consistency of the DMIP-2 QPE data caused the changes in the error trend seen Fig. 2. This motivated us to analyze the consistency of DMIP-2

QPE for both NFAR and EFCR over the periods that might affect simulation bias pattern. Note that in addition to the DMIP-2 QPE, gridded air temperature data is also necessary to run a snow model component in the OHD-DM. Errors in air temperature data affect snow accumulation amounts and rates of snowmelt, which likely lead to errors in timing of streamflow. Such errors were not as apparent in initial streamflow simulations; therefore in this study we focus on analyses of the precipitation data.

3. Study basins and precipitation data

3.1. Study basins

The NFAR and EFCR are located in the Sierra Nevada of California and Nevada (Fig. 1). These basins are characterized by different hydrologic regimes due to their mean elevation and locations on either side of the Sierran divide (Simpson et al., 2004).

The NFAR (drainage area 886 km²) is located on the western, windward side of the Sierran divide. Due to its large elevation range (260–2600 m), precipitation is enhanced by orographic effects, with mean annual precipitation varying from 813 mm at Auburn (elev. 393 m. above m.s.l., ID 37 in Fig. 1) to 1651 mm at Blue Canyon (elev. 1676 m. above m.s.l., ID 2 in Fig. 1). Types of precipitation can be rain, snow, or a mixture of rain and snow depending on the altitude and types of storms (Lundquist et al., 2008).

On the other hand, the EFCR (drainage area 714 km²) is a high elevation basin, ranging from 1650 m to about 3400 m at the basin divide. The EFCR generally drains from south to north. Due to its high elevation, most of the precipitation occurs as snowfall.

3.2. Precipitation data

To evaluate the DMIP-2 QPE data, we used the monthly precipitation data from the Parameter-Elevation Regressions on Independent Slopes Model (PRISM; Daly et al., 1994) as a reference gridded precipitation dataset for comparison. The following subsections provide brief descriptions of gridding methodologies and source gauge data for these two datasets.

3.2.1. Monthly PRISM precipitation data

Detailed discussions of the PRISM interpolation techniques are beyond the scope of this paper, but can be found in several publications (e.g., Daly et al., 2002, 2008, 1994). In this paper, a brief overview is provided instead.

One of the PRISM products (<http://www.prism.oregonstate.edu/>) is a monthly total precipitation accumulation at each 2.5-min (approximately 4 km) grid over the conterminous US, derived using linear regression between the gauge measurements and the elevation of the gauge taken from a digital elevation model (DEM). The monthly PRISM data is available from 1895 through current year. The PRISM technique attempts to account for topographic effects such as upslope precipitation enhancement and rain shadows on climate variables. The gauge measured precipitation data are supplied by various sources including the NWS Cooperative Observer Network (COOP), and the Natural Resources Conservation Service (NRCS) daily snow pack telemetry (SNOTEL) gauges. The precipitation–elevation regression equations are developed for each individual pixel using the data from gauges selected by a search window. PRISM also defines the topographic orientation derived from the digital elevation model (DEM) and includes only gauges located on slopes with the same orientation to account for local orographic effects. It should be noted that the PRISM methodology changed in the beginning of 1998. From 1895 to 1997 monthly climate data were produced retrospectively based on a 4 km DEM (Gibson et al., 2002). From 1998 to the

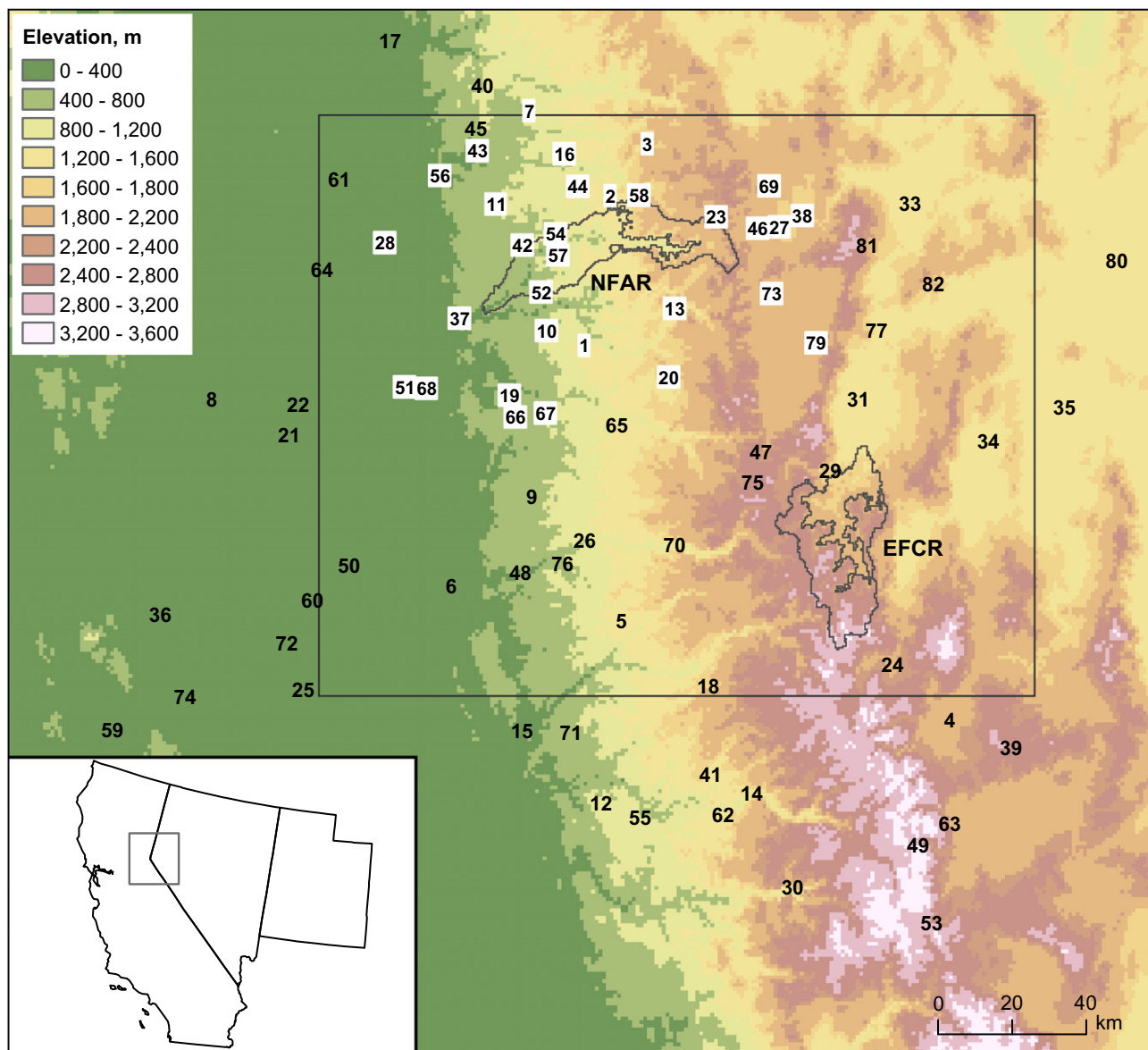


Fig. 1. North Fork American River basin and East Fork Carson River basin (upper and lower zones separated by 1524 meter contour for NFAR and by 2134 meter contour for EFCR) and NCDC and SNOTEL rain gauges used to generate DMIP-2 gridded QPE. DMIP2 QPE grid domain is indicated in a box. The numbers indicates the location of the NCDC gauges. Letter "S" indicates the location of NRCS SNOTEL site. The gauges highlighted in white are within 50 km from the NFAR boundary and listed in Fig. 3 for their gauge record history.

present, however, PRISM 1961–1990 mean monthly precipitation grids were used as a predictor in place of the DEM for the linear regression model (Wayne Gibson, PRISM Climate Group, Oregon State University, personal communication).

3.2.2. Initial 1-h DMIP-2 QPE data

The method for deriving the initial 1-h DMIP-2 gridded QPE was based on an inverse distance interpolation technique (Moreda et al., 2006; Schaake et al., 2004) to estimate precipitation at each 4 km Hydrologic Rainfall Analysis Project cell (HRAP; Reed and Maidment, 1999), which is a grid coordinate system used operationally for NWS's river forecasting. In this method, the ratios of hourly gauge precipitation to the corresponding PRISM 30-year monthly climatology (1961–1990) were interpolated instead of actual precipitation amount measured at the gauge. This approach attempts to account for climatological spatial patterns of precipitation from PRISM climatology grids during the corresponding

month, but the interpolated values are predominantly affected by hourly gauge measurements. One-hour precipitation estimates at each grid box are computed by multiplying interpolated fractions of PRISM climatology by the corresponding PRISM pixel value. In addition, the optimal power of the distance between interpolating point and gauges for the weight was identified based on minimum root mean square error of precipitation estimates. Gauge data sources for the period from 1987 through 2006 are NRCS SNOTEL and NWS COOP daily and hourly gauges, which are common gauge data sources for PRISM data. Fig. 1 shows the DMIP QPE grid domain and the spatial distribution of all the gauges used for the DMIP-2 QPE generation around the NFAR and EFCR. At an hourly time step, the number of gauges that is used for interpolation varies due to missing data and discontinued or new gauge installations. Fig. 3 presents the history of each precipitation gauge within 50 km of the NFAR basin boundary. In Fig. 3, each square represents a month in which observations were collected more

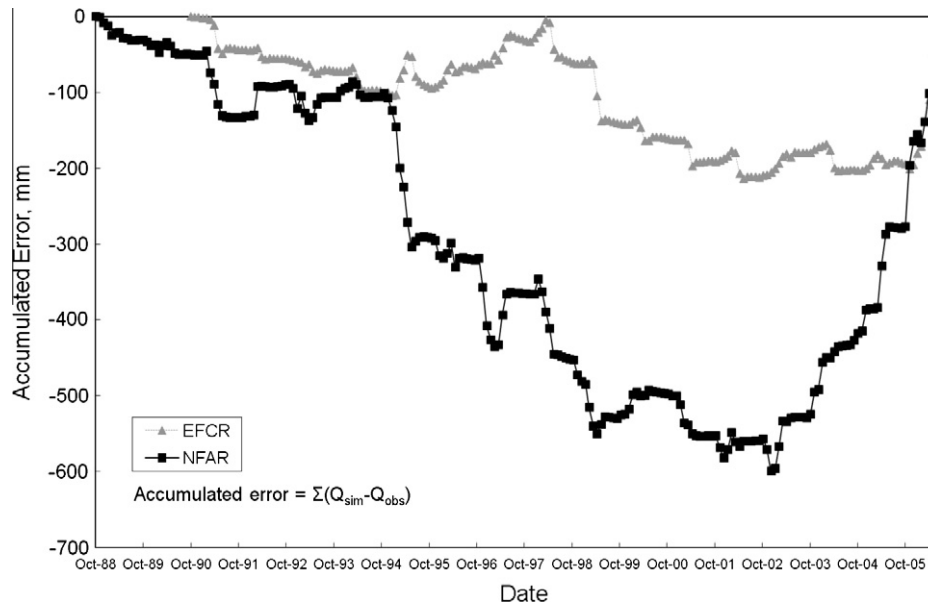


Fig. 2. Time series of cumulative error in OHD-DM simulated discharge at outlet of NFAR from WY-1988 to WY-2006 and EFCR from WY-1989 to WY-2006.

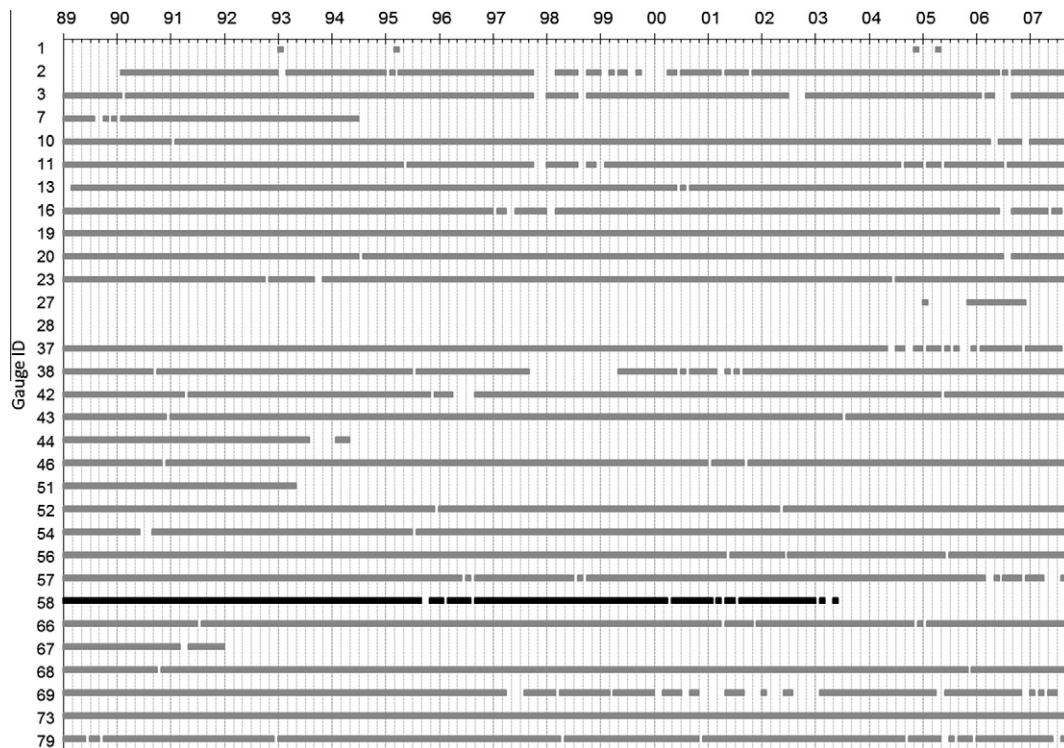


Fig. 3. Observation history for NCDC rain gauges surrounding NFAR. Y-axis is station numeric ID corresponding to the gauge ID in Fig. 1. Square symbols indicate months in which more than 80% of hourly or daily data in one month were reported as valid data for that month. The history for Gauge ID 58 – Lake Spaulding is denoted by black to highlight the measurement discontinuity. Labels on X-axis indicate January of each year.

than 80% of the time (i.e., more than 80% non-missing data). The history of individual precipitation stations near the NFAR reveals that several gauges (Gauge IDs 7, 43, 53 and 61) were discontinued during the analysis period whereas some other gauges had been reporting intermittently over the analysis period. We identified the Lake Spaulding COOP gauge (ID 58, NCDC gauge 04-4713) as being highly suspect, given its location just north of the NFAR boundary (see Fig. 1). Moreover, the Lake Spaulding gauge was discontinued in February 2003, corresponding to the abrupt change in

the streamflow error trend shown in Fig. 2. This time-varying gauge network over the simulation period likely produced inconsistent gridded QPE data, which will be uncovered using a few consistency analyses presented in the following section.

4. Consistency analysis for gridded precipitation data

We performed various analyses to evaluate the consistency of the DMIP-2 QPE. These consist of simple comparisons with an

independent gridded dataset that is commonly available to the hydrologic community (i.e. PRISM), double mass analyses (DMA), and a statistical test. Unlike many climatological studies that investigated annual time series data at meteorological stations (e.g., Alexandersson, 1986; Lund and Reeves, 2002; Rasmussen, 2001), our study analyzed gridded data on a monthly basis to correspond to the monthly PRISM data used herein as a reference.

4.1. Comparison of MAP time series

First, monthly basin-wide mean areal precipitation (MAP) time series were computed from hourly DMIP-2 QPE and monthly PRISM precipitation grids to make an overall assessment. As mentioned earlier, both datasets have similar grid sizes (~ 4 km) and use the same gauge data sources (NWS COOP and NRCS SNOTEL). However, the methodology of interpolation is different; therefore it is likely that there is a consistent difference in precipitation estimates. This analysis is not intended to prove which dataset is more accurate, but rather to evaluate the consistency of the two derived MAP time series. To investigate this, cumulative PRISM MAP values were subtracted from cumulative DMIP-2 MAP values at each monthly time step (Fig. 4). As shown in the plot, the difference between the two datasets is consistent over the EFCR (i.e. the DMIP-2 MAP is consistently greater than PRISM MAP), indicated by the approximately constant upward slope in the plot. On the other hand, for NFAR, the DMIP-2 MAP produces an underestimation until the end of 2002 and then switches to overestimation. In addition, a discontinuity is seen around January 1995, which is coincident with the trend in cumulative error in streamflow simulation as seen in Fig. 2. We repeated this MAP analysis with two elevation zones partitioned at the 1524 m contour line. These zones are used for lumped model operational river forecasting by the NWS, California Nevada River Forecast Center (CNRFC). The same consistency change in the difference of the two MAP time series was seen over the lower zone and upper zone in NFAR (not shown here).

Based on these results, we conclude that there is no major inconsistency in the DMIP-2 gridded QPE for EFCR relative to PRISM data. However, it is highly likely that DMIP-2 QPE for NFAR does contain inconsistencies, leading to the trends in streamflow simulation seen in Fig. 2. Therefore the analyses and discussions in the rest of the paper are focused on the NFAR domain.

4.2. Double mass analysis

In order to evaluate the inconsistencies of the DMIP-2 QPE independently from PRISM, which might also contain inconsistencies due to the methodology change in 1997, we performed a double mass analysis (DMA) for a quick visual check of inconsistencies in the time series. DMA is a graphical tool traditionally used to detect inconsistencies in multi-year gauge data by comparing cumulative time series at one gauge to a 'reference' cumulative time series based on several nearby gauges (Chang and Lee, 1974; Kohler, 1949; Searcy and Hardison, 1960). Bends or deflections in the double mass curve reflect changes in gauge catch characteristics that produce a data inconsistency at the suspected gauge. In contrast to a single gauge whose data acquisition is independent from the other gauges, an individual grid cell shares multiple gauges with other grid cells because of the spatial interpolation process. In other words, multiple neighboring cells may contain similar inconsistencies. However, the degree of inconsistency should vary from cell to cell. For gridded precipitation data generated based on an inverse distance technique like the DMIP-2 QPE, for example, cells that are farther from a suspect gauge are less affected than closer ones. To address this, we formed eight cell groups, each of which consists of four cells that might have a similar degree of inconsistency, to perform the double mass analysis for each cell group as shown in Fig. 5.

For DMA, generating a consistent reference series is crucial for detecting inconsistencies in a series at a target location (i.e. cell group). In this study, four neighboring gauges were selected around NFAR (Gauge-IDs 13, 23, 52 and 54 in Figs. 1 and 3) to generate the reference time series. These gauges provide complete time series of daily or hourly precipitation values from WY-1989 to WY-2006. Each gauge time series was quality controlled by following the techniques used for NWS historical data analysis (Anderson, 2002; Smith et al., 2003) and then aggregated to monthly totals. The following formula (Beaulieu et al., 2008) was used to generate a monthly reference series R_i for each target cell group

$$R_i = \frac{\sum_{j=1}^4 \rho_j^2 \frac{p}{G_j} G_{ij}}{\sum_{j=1}^4 \rho_j^2} \quad i = 1, \dots, n \quad (1)$$

where n is the number of months from October 1988 to September 2006, ρ_j is the correlation coefficient between monthly time series

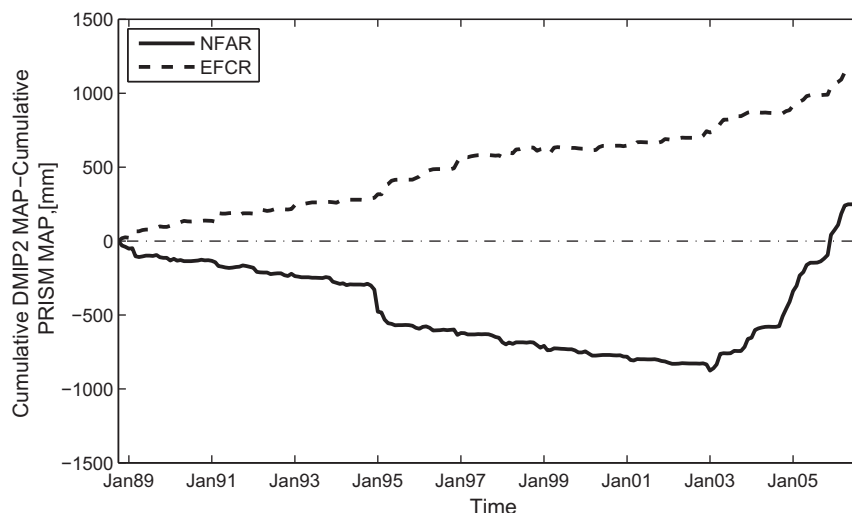


Fig. 4. Cumulative difference between monthly MAP from PRISM and MAP from DMIP-2 gridded QPE for EFCR and NFAR.

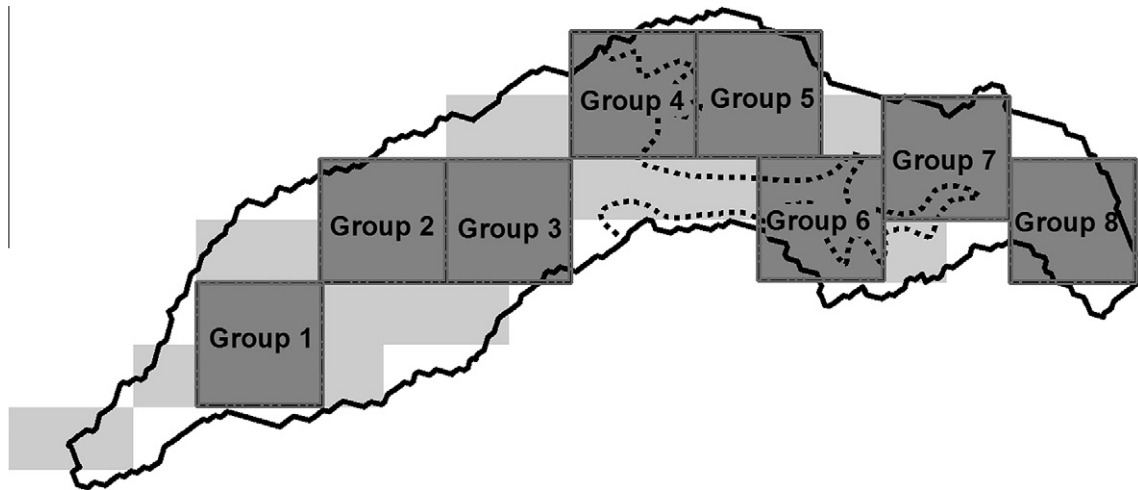


Fig. 5. NFAR basin defined on 4 km HRAP grid, indicated by light gray cells, and eight cell groups used for double mass analysis. Each cell group consists of four HRAP grid cells. NFAR upper and lower elevation zones separated by 1524 m contour are also shown.

at a target cell group and at a neighbor gauge series j , G_{ij} is the monthly series of a neighbor gauge j , \bar{P} is the mean of the monthly series at the target cell group, and \bar{G}_j is mean of the monthly series at the neighbor gauge j .

Precipitation in the Sierra Nevada exhibits strong seasonal patterns (e.g., Dettinger et al., 2004). Much less precipitation occurs during the summer season than during the winter season, and zero precipitation is recorded during some summer months at some low elevation cells. We used only the data during winter months (October through May) because the inconsistency in the time series appears more clearly during the winter due to the greater amount of precipitation and subsequent influence on the hydrologic processes over the study area.

In this DMA, the cumulative MAP time series of each cell group was plotted against the cumulative reference series. The plots for DMIP-2 QPE and PRISM are shown in Figs. 6 and 7, respectively. To highlight any breaks in the slope, we plot the deviation of the accumulated four-cell group precipitation from the accumulation of the reference data as is done in NWS historical data analysis (Anderson, 2002; Smith et al., 2003). A constant slope of the double mass curve indicates that the accumulation rate of a target cell group is consistent compared to the reference series. A change in the slope indicates a change in accumulation rate of a target cell group relative to the reference data, thus an inconsistency. In the DMA, each point in the curve corresponds to a particular time of the series (i.e. month). A visual inspection of the DMIP-2 QPE double mass curve reveals that a clear slope deflection occurs in higher elevation cell groups (cell groups – four through eight) in January 2003. This break in the slope indicates a severe inconsistency of the DMIP-2 QPE data: from 1988 to 2003, the accumulation slope is negative, while after 2003 the slope is positive compared to the reference series. This inconsistency is less severe for cell group three compared to the higher elevation cell groups, and there is no clear inconsistency found in the time series for cell groups one and two. Based on the timing of this deflection, it is concluded that the bias pattern change in streamflow simulation seen in Fig. 2 is associated with the DMIP-2 QPE inconsistency.

Fig. 7 shows there is also some inconsistency in PRISM data. However, the PRISM data exhibits a different type of inconsistency than in the DMIP-2 QPE data. For all the cell groups except cell group one, the accumulation rate is approximately equal to the rate for the corresponding reference series until a certain time. Thereafter, the accumulation rate becomes less compared to the reference series. The timing and manner of this pattern changes

from cell group to cell group. It does not appear that the breaks in double mass curves reflect the change in the PRISM technique in 1997.

4.3. Statistical tests

For more objective detection of inconsistencies in precipitation and other hydrometeorological time series, there are various statistical tests available. Several authors have presented comparative studies on the performance of statistical tests (e.g., Beaulieu et al., 2008; Buishand, 1982; Ducré-Robitaille et al., 2003; Sahin and Cigizoglu, 2010). Statistical methods include *relative* tests such as the Bivariate test, which like DMA rely on a reference time series, and *absolute* tests including the Standard Normal Homogeneity Test (SNHT), the Sequential T test, the Sequential Wilcoxon test and the Bayesian method, which do not require a reference dataset. Many of these techniques are based on assumptions such as data normality, which is an obstacle in our case since monthly precipitation data do not follow a normal distribution (Beaulieu et al., 2008). It is noted that most of the past studies about climate data consistency used annual time step datasets, which are assumed to be normally distributed (e.g., Ducré-Robitaille et al., 2003; Lund and Reeves, 2002).

Following the SNHT test procedure demonstrated by Beaulieu et al. (2008), we computed a time series of the ratio q_i between the target and reference series on a monthly basis. This improves the normality of the time series data. A χ^2 square goodness of fit test (Wilks, 2005) was performed on the monthly ratio time series of DMIP-2 QPE for each cell group to assess the normality (Table 1). The p -value of the χ^2 square test statistic computed using the ratio time indicate the ratio series for all the cell group exhibits normality at 90% significance (the p -value greater than 0.1) while monthly precipitation time series do not fit a normal distribution well.

It is possible to perform several *absolute* tests given the normality of the ratio series. The objective of this study is not to evaluate the performance of various statistical tests for the gridded dataset used in this study. Rather, in this study, we use SNHT because of its simplicity and relatively good performance compared to other tests (Beaulieu et al., 2008). In SNHT, the null hypothesis that a standardized anomaly time series (in this case, ratio series q_i is standardized) follows a normal distribution with zero mean and a standard deviation of one is tested against the alternative hypothesis that there is shift in the mean. Using the ratio series, the weighted average series of the ratio T_i is defined by,

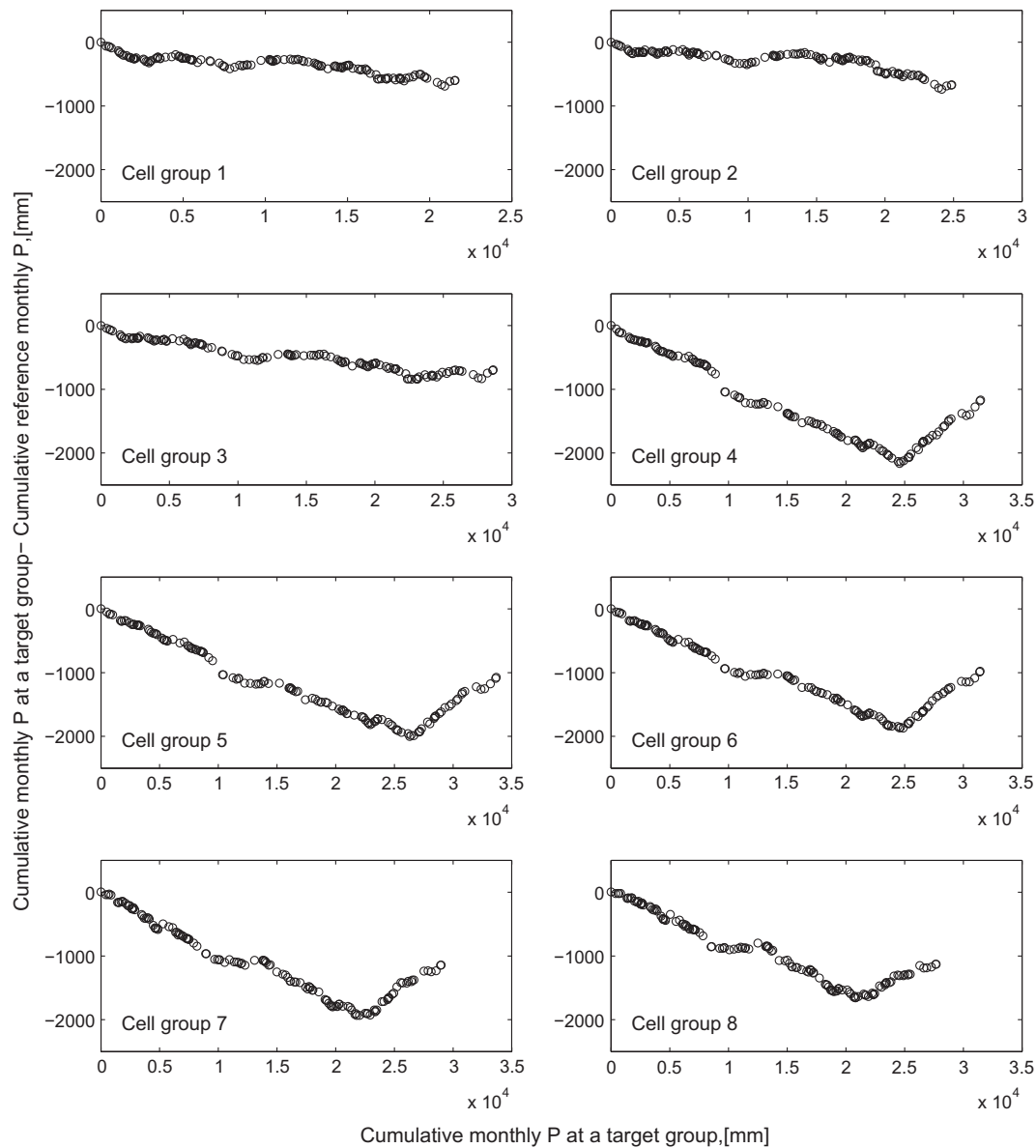


Fig. 6. Double mass curves of cell group DMIP-2 monthly series against a reference series for eight cell groups. X-axis is the accumulation of the reference series; Y-axis shows the deviation of the accumulation of the subject four-cell group.

$$T_i = i \cdot \bar{z}_1^2 + (n - i) \cdot \bar{z}_2^2 \quad i = 1, \dots, n \quad (2)$$

$$\bar{z}_1 = \frac{1}{i} \sum_{j=1}^i (q_j - \bar{q}) / s_q$$

$$\bar{z}_2 = \frac{1}{n - i} \sum_{j=i+1}^n (q_j - \bar{q}) / s_q$$

where \bar{q} is mean of the ratio series, n is the sample size (in this case the number of time steps or 144), and s_q is its standard deviation. The test statistic T_0 is defined as the maximum $\{T_i\}$. If T_0 exceeds the critical value, which depends on sample size (Khaliq and Ouada, 2007), the null hypothesis will be rejected. Although DMA methods graphically indicated only one abrupt change in the error consistency in the time series, examination of the time series of the T statistic can quantify the degree of abrupt changes in the error consistencies at each time step throughout the time series.

The results of SNHT for the DMIP-2 QPE are shown in Fig. 8. T_0 values for cell groups four, five, six and seven exceed the 99% level

critical value of 12.6 for the sample size n of 144. The timing of T_0 also corresponds to the slope break point in the double mass curves (Fig. 6). These results indicate a statistically significant inconsistency for these data around 2003. T_0 values for cell groups one, two, three, and eight also peak around beginning of 2003, although the values do not exceed the critical value. Unlike the T values from the DMIP-2 QPE data, T values computed with PRISM data exhibit much less distinct peaks in the time series (not shown). This is because the inconsistency in the PRISM data occurs more gradually as seen in the double mass curves of Fig. 7. Therefore the timing of T_0 for the PRISM data is also difficult to determine.

5. Influence of precipitation data consistency on hydrologic simulation

To stress the importance of precipitation data consistency for the simulation of streamflow, hydrologic simulations were performed in a lumped mode using two MAP datasets: the DMIP-2

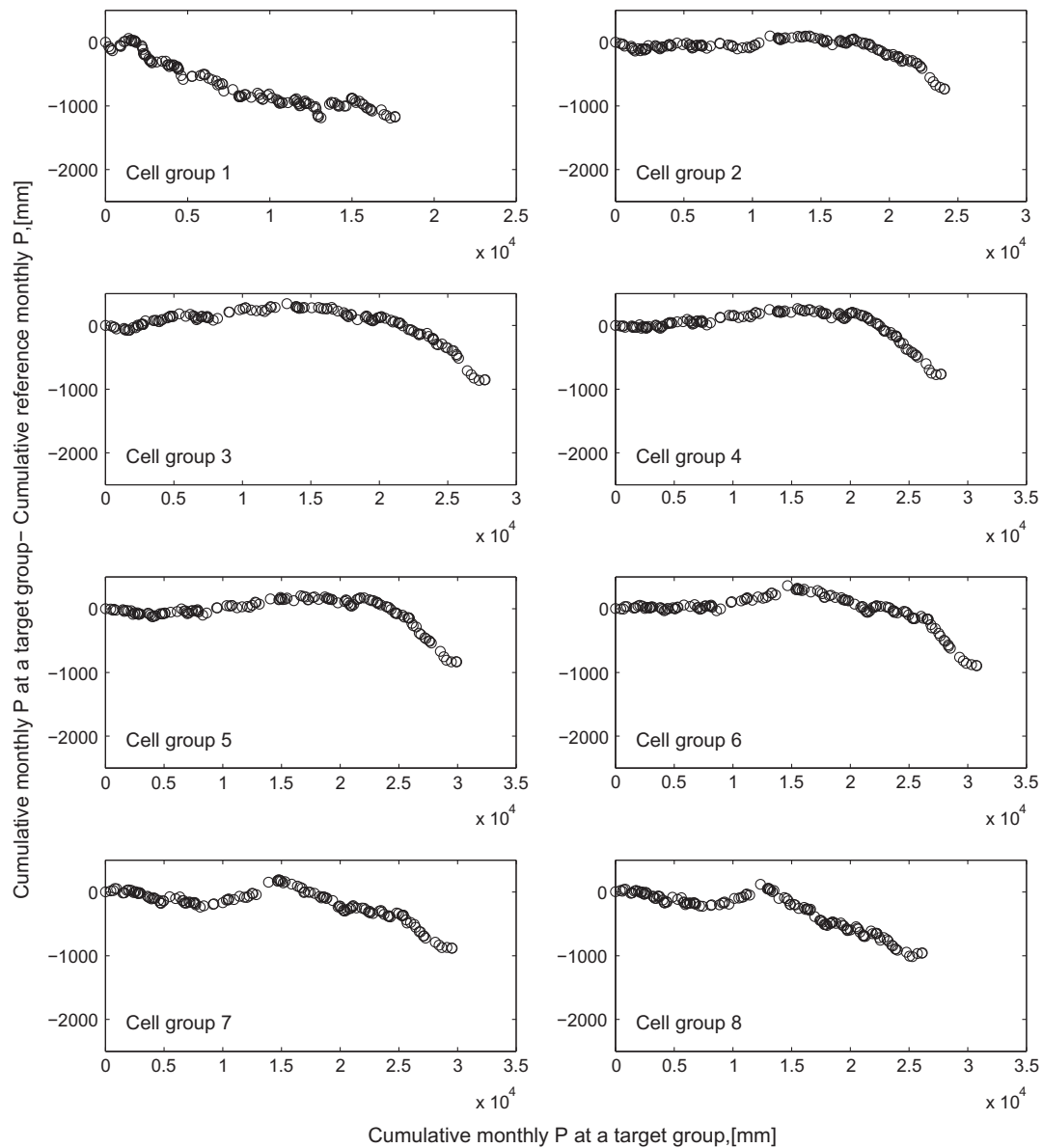


Fig. 7. The same as Fig. 6 except using PRISM monthly series.

Table 1

The p -value of test statistics obtained from χ^2 distribution for monthly precipitation time series and time series of the ratio q_i between monthly precipitation at cell group and reference gage precipitation R (Eq. (1)). The higher p -value indicates less likelihood that null hypothesis that the time series data are drawn from normal distribution is rejected.

	Cell group 1	Cell group 2	Cell group 3	Cell group 4	Cell group 5	Cell group 6	Cell group 7	Cell group 8
Precipitation	0.041	0.134	0.068	0.015	0.002	0.002	0.009	0.004
Ratio, q_i	0.659	0.130	0.529	0.183	0.328	0.453	0.756	0.364

MAP mentioned in Section 4, which contains inconsistent bias and an 'adjusted' DMIP-2 MAP, in which the inconsistency was corrected. Zhang et al. (2010) described a method for correcting radar-based QPE bias on a cell-by-cell basis using PRISM monthly precipitation estimates. Since the PRISM estimates are subject to some inconsistencies for our study area shown in Fig. 7, we decided to forgo the PRISM-based adjustments. Instead, 1-h DMIP-2 MAP time series defined for the two NFAR elevation zones were adjusted so that the double mass slope (i.e. ratio of DMIP-2 MAP accumulation rate to the rate for the reference data) from

January 2003 onward matches the accumulation rate between October 1988 and January 2003. This double mass-based correction could be done per an individual grid cell. However, our goal here was not to correct the inconsistency of the DMIP-2 QPE gridded data, but to illustrate the effect of data consistencies on overall simulation results. In fact, we developed and successfully used an entirely new method to derive gridded QPE data for the DMIP 2 Sierra Nevada basins (Smith et al., 2010). Our experience with lumped and distributed modeling in NFAR has shown that inconsistent QPE data leads to similar behavior in accumulated runoff

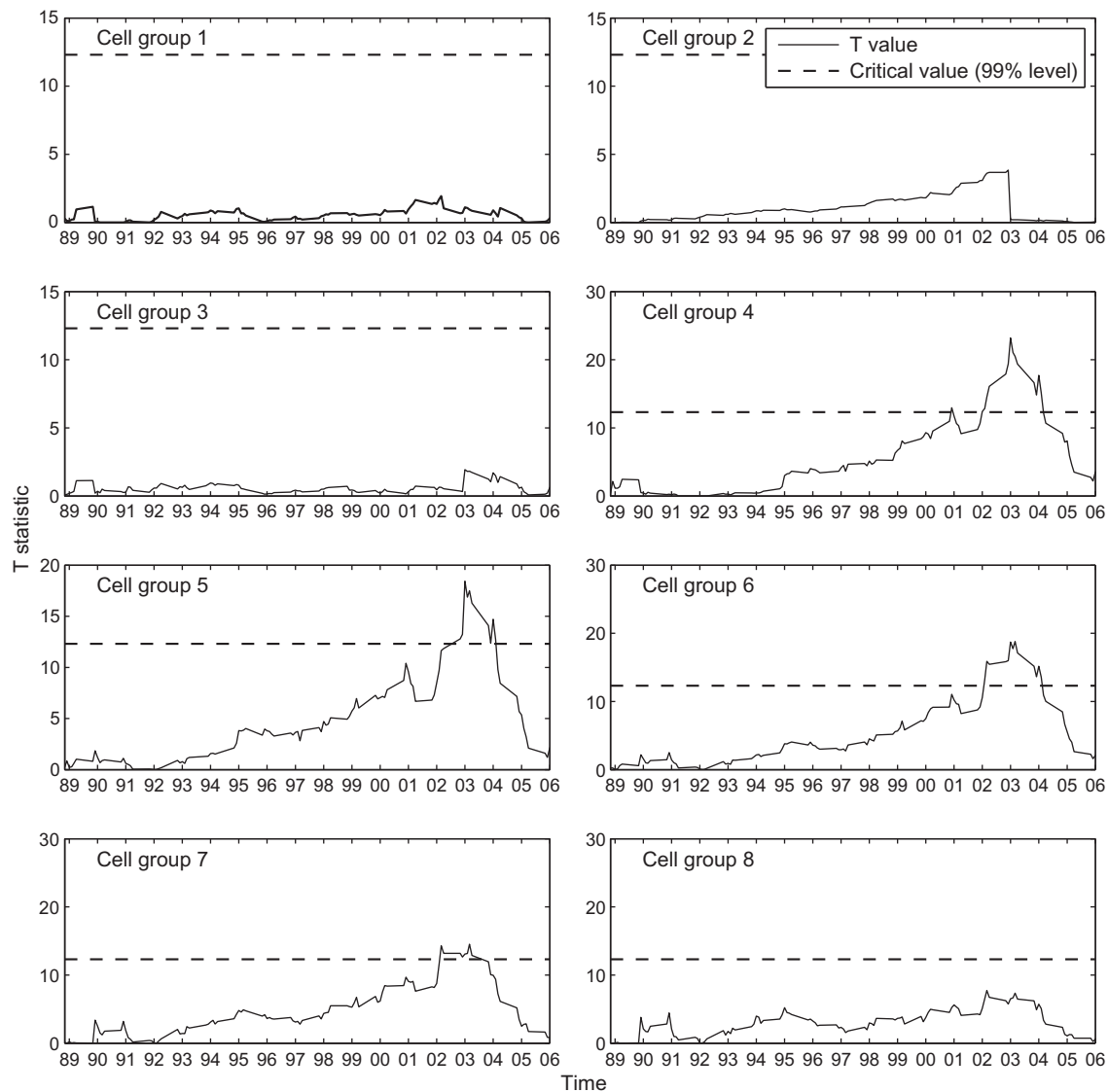


Fig. 8. Time series of SNHT T values computed using the four-cell group DMIP-2 monthly series for each cell group. Tick marks on X-axis indicate January of each year. Dashed line indicates the critical T_0 value of 12.6.

error for both modeling approaches. Therefore for simplicity, the lumped model was chosen over the distributed model for this illustration.

Fig. 9 shows the double mass curve for the zonal MAPs computed with gridded DMIP-2 data as well as PRISM data. The same types of inconsistencies as in the finer spatial scale MAP series (i.e. MAP for eight cell groups) are observed for the DMIP-2 and PRISM MAPs. The adjustment factor (AF) to correct the DMIP-2 MAP inconsistency was computed as follows:

$$AF = \frac{B_{1:m}}{B_{m:n}} \quad \text{where } B_{s:t} = \frac{\sum_{i=1}^t P_i - \sum_{i=1}^s P_i}{\sum_{i=1}^t R_i - \sum_{i=1}^s R_i} \quad (3)$$

where $B_{s:t}$ is a slope of a double mass curve between s th and t th months, m is a month when a double mass break occurs, P_i is monthly total at a target location (i.e. zonal MAP) during the i th month and R_i is a monthly value in a reference series (see Eq. (1)) during i th month. Each 1-h DMIP-2 MAP value after the m th month was multiplied by the value of AF. January 2003 is identified as the month containing the double mass break and AF values were computed for the lower and upper zones (0.82 and 0.92, respectively).

The lumped simulations are derived using the same snow and rainfall-runoff models as in the OHD-DM. The lumped model is run over the two elevation zones of NFAR instead of at each HRAP grid cell. The lumped model also requires a unit hydrograph to transform runoff depth to discharge whereas the OHD-DM utilizes a kinematic hillslope and channel routing scheme to route runoff volumes. Despite the difference in routing methods, the impact of inconsistent precipitation data on the simulation is presumed to be the same based on our past experience. The lumped model parameters used here were calibrated with 1-h DMIP-2 MAPs for the upper and lower zone as well as zonal mean areal temperature (MAT) data, which was also computed from 1-h gridded air temperature data generated for DMIP-2. The calibration period for the lumped models is from October 1988 through September 1997—therefore the model calibration was not affected by the data inconsistency identified in January 2003.

Fig. 10 shows the time series of cumulative errors from the two lumped simulations, which were made with the same model parameters. As seen in Fig. 10, the simulation with the original DMIP-2 MAP is relatively stable before January 2003 except for the sharp drop in the beginning of 1995, which was also observed

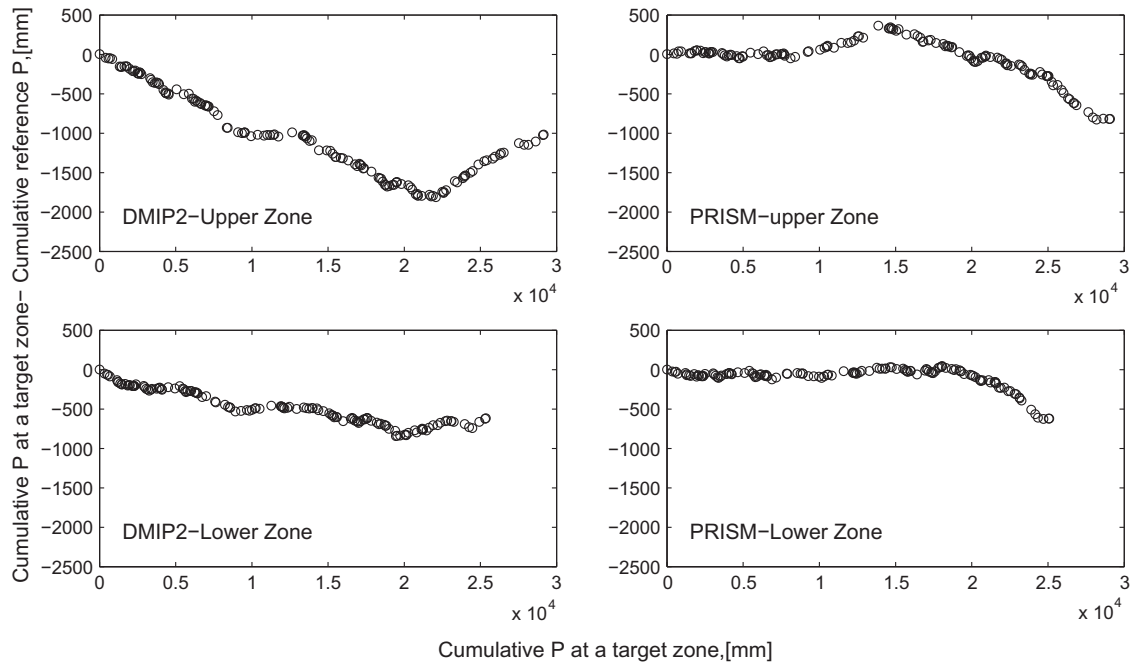


Fig. 9. Double mass curve for DMIP-2 MAP (left column) and PRISM MAP (right column) for NFAR upper (top row) and lower zone (bottom row).

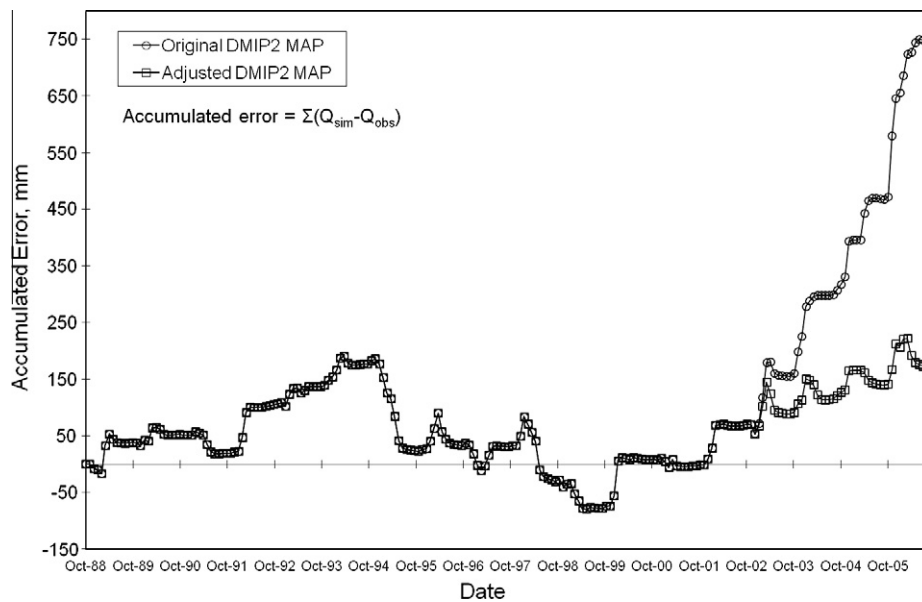


Fig. 10. Monthly accumulated error of 1-h lumped model simulations for NFAR using (a) 1-h MAP created using original DMIP-2 QPE and (b) 1-h MAP adjusted based on double mass curve with the same model parameters. The lumped model was calibrated with lower and upper zone MAP during WY-1989 through WY-1996 using the original DMIP-2 MAP.

in the OHD-DM simulation seen in Fig. 2. However, the simulation starts overestimating after January 2003, corresponding to the occurrence of the DMIP-2 QPE inconsistency. This drastic increase in error is the same behavior seen in Fig. 2 in the result from the OHD-DM. By contrast, an abrupt increase in error trend after January 2003 was not shown in the streamflow simulation with the corrected DMIP-2 MAPs (the second simulation).

6. Conclusions

Consistent long term QPE is important to simulation with hydrologic models such as the ones used operationally in NWS

River Forecast Centers. Given the expanding operational use of distributed models, spatially distributed QPE time series for mountainous basins are essential. At least for now, such QPE products are based in large part on point gauge measurements.

We presented an approach to uncover the inconsistencies potentially contained in multi-year gridded datasets. We investigated the hourly gridded QPE from WY-1988 through 2006 generated via interpolation of gauge network data over the mountainous areas. Traditional DMA methods were applied to gridded data to identify inconsistencies. This simple technique is shown to be capable of detecting inconsistencies in gridded datasets at grid cell locations. In addition to DMA, one of the statistical tests, SNHT, was

performed on the DMIP-2 QPE, showing a clear inconsistency in the same cell groups in the dataset and occurring at the same timing as indicated by DMA. Potentially, such statistical tests as well as DMA could be used to detect inconsistencies on a cell by cell basis objectively. Finally in order to illustrate the impact of input data consistency on hydrologic simulations, two streamflow simulations were made using the inconsistent and inconsistency-corrected MAP series. Not surprisingly, streamflow simulations based on the corrected input QPE data are more consistent and accurate than those based on the inconsistent QPE data. Although a lumped model was used for this discussion, the results and conclusions are assumed to be transferable to distributed hydrologic model simulation with gridded QPE. The methodology of adjustment for gridded precipitation data over complex terrain on a cell-by-cell basis still needs to be investigated.

At a minimum, the generation of consistent gauge-based QPE grids requires following throughout the study period – (1) gauge observations that are consistent, (2) a gauge network that is stable, and (3) an effective technique to interpolate grid values from point observations. One of the plausible causes of the inconsistency in the initial DMIP-2 QPE is the discontinuation of one of the gauges closest to the NFAR. A possible solution to this problem is to generate complete and consistent time series at each station for the analysis period using historical data analysis such as the methods described in Smith et al. (2003) and Eischeid et al. (2000). These complete and consistent station time series could then be spatially interpolated to form a gridded QPE dataset.

This quality control of gauge data is essential to the generation of long term consistent radar-based QPE as well. As in non-mountainous areas, radar-based QPE relies partly on gauge data to adjust radar-only QPE to generate multisensor precipitation estimates (Seo, 1998). While radar-only QPE fields can possess more reasonable spatial variability than gauge-only gridded products, caution should be taken in merging radar QPE and gauge data over multi-year periods. As experienced in DMIP-2, where hourly gauge data was interpolated over the areas continuously for 20 years, frequent gauge network changes due to missing data, gauge relocation and discontinuation could adversely affect the radar based QPE. In addition, there are several studies to develop multi-year gridded datasets based on reanalyses using consistent hydrometeorological model simulations and other means (e.g., Nelson et al., 2010). Temporally consistent forcing grids as well as the same model configuration throughout the modeling period are necessary to generate consistent hydrometeorological gridded outputs.

Acknowledgements

The authors thank the reviewers for their comments and suggestions that improved the quality of this manuscript.

References

- Adam, J.C., Lettenmaier, D.P., 2003. Adjustment of global gridded precipitation for systematic bias. *Journal of Geophysical Research* 108 (D9), 4257.
- Alexanderson, H., 1986. A homogeneity test applied to precipitation data. *Journal of Climatology* 6 (6), 661–675.
- Anderson, E.A., 2002. Calibration of Conceptual Models for Use in Operational River Forecasting. NOAA NWS Office of Hydrologic Development. <http://www.nws.noaa.gov/oh/hrl/calb/calibration1102/calb_report_contents_etc.pdf>.
- Andréassian, V., Perrin, C., Michel, C., Usart-Sanchez, I., Lavabre, J., 2001. Impact of imperfect rainfall knowledge on the efficiency and the parameters of watershed models. *Journal of Hydrology* 250 (1–4), 206–223.
- Bales, R.C. et al., 2006. Mountain hydrology of the western United States. *Water Resources Research* 42 (8), W08432.
- Beaulieu, C. et al., 2008. Intercomparison of homogenization techniques for precipitation data. *Water Resources Research* 44 (2), W02425.
- Brath, A., Montanari, A., Toth, E., 2004. Analysis of the effects of different scenarios of historical data availability on the calibration of a spatially-distributed hydrological model. *Journal of Hydrology* 291 (3–4), 232–253.
- Buishand, T.A., 1982. Some methods for testing the homogeneity of rainfall records. *Journal of Hydrology* 58 (1–2), 11–27.
- Chang, M., Lee, R., 1974. Objective double mass analysis. *Water Resources Research* 10 (6), 1123–1126.
- Crum, T.D., Alberty, R.L., 1993. The WSR-88D and the WSR-88D operational support facility. *Bulletin of the American Meteorological Society* 74 (9), 1669–1687.
- Daly, C., Neilson, R.P., Phillips, D.L., 1994. A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *Journal of Applied Meteorology* 33 (2), 140–158.
- Daly, C., Gibson, W.P., Taylor, G.H., Johnson, G.L., Pasteris, P., 2002. A knowledge-based approach to the statistical mapping of climate. *Climate Research* 22 (2), 99–113.
- Daly, C. et al., 2008. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology* 28 (15), 2031–2064.
- Dawdy, D.R., Bergmann, J.M., 1969. Effect of rainfall variability on streamflow simulation. *Water Resources Research* 5 (5), 958–966.
- Dettinger, M.D., Cayan, D.R., Meyer, M.K., Jeton, A.E., 2004. Simulated hydrologic responses to climate variations and change in the Merced, Carson, and American River basins, Sierra Nevada, California, 1900–1999. *Climate Change* 62, 283–317.
- Ducré-Robitaille, J.-F., Vincent, L.A., Boulet, G., 2003. Comparison of techniques for detection of discontinuities in temperature series. *International Journal of Climatology* 23 (9), 1087–1101.
- Eischeid, J.K., Pasteris, P.A., Diaz, H.F., Planticaio, M.S., Lott, N.J., 2000. Creating a serially complete, national daily time series of temperature and precipitation for the western United States. *Journal of Applied Meteorology* 39, 1580–1591.
- Fulton, R.A., Breidenbach, J.P., Seo, D.-J., Miller, D.A., O'Bannon, T., 1998. The WSR-88D rainfall algorithm. *Weather and Forecasting* 13 (2), 377–395.
- Gan, T., Dlamini, E.M., Biftu, G.F., 1997. Effects of model complexity and structure, data quality, and objective functions on hydrologic modeling. *Journal of Hydrology* 192 (1–4), 81–103.
- Gibson, W.P., et al., 2002. Development of a 103-year high-resolution climate data set for the conterminous United States. In: 13th Conference of Applied Climatology, American Meteorological Society Meeting, Portland, OR.
- Gourley, J.J., Jorgensen, D.P., Matrosov, S.Y., Flamig, Z.L., 2009. Evaluation of incremental improvements to quantitative precipitation estimates in complex terrain. *Journal of Hydrometeorology* 10 (6), 1507–1520.
- Guentchev, G., Barsugli, J.J., Eischeid, J., 2010. Homogeneity of gridded precipitation datasets for the Colorado River Basin. *Journal of Applied Meteorology and Climatology* 49 (12), 2404–2415.
- Hamlet, A.F., Lettenmaier, D.P., 2005. Production of temporally consistent gridded precipitation and temperature fields for the continental United States. *Journal of Hydrometeorology* 6 (3), 330–336.
- Khaliq, M.N., Ouarda, T.B.M.J., 2007. On the critical values of the standard normal homogeneity test (SNHT). *International Journal of Climatology* 27 (5), 681–687.
- Kohler, M.A., 1949. Double-mass analysis for testing the consistency of records and for making required adjustments. *Bulletin of the American Meteorological Society* 30, 188–189.
- Koren, V., Reed, S., Smith, M., Zhang, Z., Seo, D.-J., 2004. Hydrology laboratory research modeling system (HL-RMS) of the US national weather service. *Journal of Hydrology* 291 (3–4), 297–318.
- Legates, D.R., 1987. A climatology of global precipitation. *Publications in climatology* 40 (1), 86.
- Lohmann, D. et al., 2004. Streamflow and water balance intercomparisons of four land surface models in the North American Land Data Assimilation System project. *Journal of Geophysical Research* 109 (D7), D07S91.
- Lund, R., Reeves, J., 2002. Detection of undocumented change points: a revision of the two-phase regression model. *Journal of Climate* 15 (17), 2547–2554.
- Lundquist, J.D. et al., 2008. Rain versus snow in the Sierra Nevada, California: comparing doppler profiling radar and surface observations of melting level. *Journal of Hydrometeorology* 9 (2), 194–211.
- Matrosov, S.Y., 2010. Evaluating polarimetric X-band radar rainfall estimators during HMT. *Journal of Atmospheric and Oceanic Technology* 27 (1), 122–134.
- Matrosov, S.Y., Campbell, C., Kingsmill, D., Sukovich, E., 2009. Assessing snowfall rates from X-band radar reflectivity measurements. *Journal of Atmospheric and Oceanic Technology* 26 (11), 2324–2339.
- Metcalfe, J.R., Routledge, B., Devine, K., 1997. Rainfall measurement in Canada: changing observational methods and archive adjustment procedures. *Journal of Climate* 10 (1), 92–101.
- Morela, F., Cong, S., Schaake, J., Smith, M., 2006. Gridded Rainfall Estimation for Distributed Modeling in Western Mountainous Areas. AGU Joint Assembly, Baltimore.
- Nelson, B.R., Seo, D.J., Kim, D., 2010. Multisensor precipitation reanalysis. *Journal of Hydrometeorology* 11 (3), 666–682.
- Pan, M. et al., 2003. Snow process modeling in the North American Land Data Assimilation System (NLDAS): 2. Evaluation of model simulated snow water equivalent. *Journal of Geophysical Research* 108 (D22), 8850.
- Rasmussen, P., 2001. Bayesian estimation of change points using the general linear model. *Water Resources Research* 37 (11), 2723–2731.
- Rasmussen, R. et al., 2003. Snow nowcasting using a real-time correlation of radar reflectivity with snow gauge accumulation. *Journal of Applied Meteorology* 42 (1), 20–36.
- Reed, S., Maidment, D., 1999. Coordinate transformations for using NEXRAD DATA in GIS-based hydrologic modeling. *Journal of Hydrologic Engineering* 4 (2), 174.

- Reed, S. et al., 2004. Overall distributed model intercomparison project results. *Journal of Hydrology* 298 (1–4), 27–60.
- Sahin, S., Cigizoglu, H.K., 2010. Homogeneity analysis of Turkish meteorological data set. *Hydrological Processes* 24 (8), 981–992.
- Schaake, J., Henkel, A., Cong, S., 2004. Application of PRISM climatologies for hydrologic modeling and forecasting in the western US. In: 18th Conference of Hydrology, 84th American Meteorological Society Meeting, Seattle, WA.
- Searcy, J.K., Hardison, C.H., 1960. Double-Mass Curves, Manual of Hydrology: Part 1, General Surface Water Techniques. United States Government Printing Office, Washington, DC, pp. 31–66.
- Seo, D.-J., 1998. Real-time estimation of rainfall fields using radar rainfall and rain gauge data. *Journal of Hydrology* 208 (1–2), 37–52.
- Sevruk, B., 1982. Methods of correction for systematic error in point precipitation measurement for operational use Operational hydrology report, 21. Secretariat of the World Meteorological Organization, Geneva, Switzerland, 91 pp.
- Sheffield, J. et al., 2003. Snow process modeling in the North American Land Data Assimilation System (NLDAS): 1. Evaluation of model-simulated snow cover extent. *Journal of Geophysical Research* 108 (D22), 8849.
- Simpson, J.J., Dettinger, M.D., Gehrke, F., McIntire, T.J., Hufford, G.L., 2004. Hydrologic scales, cloud variability, remote sensing, and models: implications for forecasting snowmelt and streamflow. *Weather and Forecasting* 19 (2), 251–276.
- Smith, M.B., Laurine, D.P., Koren, V.I., Reed, S.M., Zhang, Z., 2003. Hydrologic model calibration in the National Weather Service. In: Duan, Q., Gupta, H.V., Sorooshian, S., Rousseau, A.N., Turcotte, R. (Eds.), *Calibration of Watershed Models*. AGU, Washington, DC, pp. 133–152.
- Smith, M.B. et al., 2004. The distributed model intercomparison project (DMIP): motivation and experiment design. *Journal of Hydrology* 298 (1–4), 4–26.
- Smith, M., Koren, V., Reed, S., Zhang, Z., Cui, Z., Mizukami, N., Moreda, F., Sheldon, S., 2009. Preliminary results of phase 2 of the distributed model intercomparison project. In: Presentation 4.1, 23rd Conference on Hydrology, 89th meeting of the AMS, Phoenix, Arizona, January 10–15. Recorded Presentation. <http://ams.confex.com/ams/89annual/techprogram/paper_148995.htm>.
- Smith, M., Ding, F., Cui, Z., Koren, V., Mizukami, N., Zhang, Z., Cosgrove, B., Kitzmiller, D., Schaake, J., 2010. Development of gridded QPE datasets for mountainous area distributed hydrologic modeling. In: EWRI World Environmental and Water Resources Congress 2010, May 16–20, 2010, Providence, Rhode Island.
- Smith, M.B. et al., 2012a. The distributed model intercomparison project – phase 2: motivation and design of the Oklahoma experiments. *Journal of Hydrology* 418–419, 3–16.
- Smith, M.B. et al., 2012b. Results of the DMIP 2 Oklahoma experiments. *Journal of Hydrology* 418–419, 17–48.
- Sorooshian, S., Gupta, H.V., 1983. Automatic calibration of conceptual rainfall-runoff models: the question of parameter observability and uniqueness. *Water Resources Research* 19 (1), 260–268.
- Sorooshian, S., Gupta, H.V., Fulton, J.L., 1983. Evaluation of maximum likelihood parameter estimation techniques for conceptual rainfall-runoff models: influence of calibration data variability and length on model credibility. *Water Resources Research* 19 (1), 251–259.
- Steiner, M., Smith, J.A., Burges, S.J., Alonso, C.V., Darden, R.W., 1999. Effect of bias adjustment and rain gauge data quality control on radar rainfall estimation. *Water Resources Research* 35 (8), 2487–2503.
- Sutcliffe, J.V., Lazenby, J.B., 1990. Hydrological Data for Surface Water Resources Assessment. The Hydrological Basis for Water Resources Management. International Association of Hydrological Sciences, pp. 251–259.
- Troutman, B.M., 1982. An analysis of input errors in precipitation-runoff models using regression with errors in the independent variables. *Water Resources Research* 18 (4), 947–964.
- Troutman, B.M., 1983. Runoff prediction errors and bias in parameter estimation induced by spatial variability of precipitation. *Water Resources Research* 19 (3), 791–810.
- Troutman, B.M., 1985. Errors and parameter estimation in precipitation-runoff modeling. 1. Theory. *Water Resources Research* 21 (8), 1195–1213.
- Warner, T.T., Brandes, E.A., Sun, J., Yates, D.N., Mueller, C.K., 2000. Prediction of a flash flood in complex terrain. Part I: A comparison of rainfall estimates from radar, and very short range rainfall simulations from a dynamic model and an automated algorithmic system. *Journal of Applied Meteorology* 39 (6), 797–814.
- Westrick, K.J., Mass, C.F., Colle, B.A., 1999. The limitations of the WSR-88D radar network for quantitative precipitation measurement over the Coastal Western United States. *Bulletin of the American Meteorological Society* 80 (11), 2289–2298.
- Wilks, D., 2005. *Statistical Methods in the Atmospheric Sciences*. Academic Press, San Diego, CA, 627 pp.
- Xie, H. et al., 2006. Evaluation of NEXRAD stage III precipitation data over a semiarid region. *JAWRA Journal of the American Water Resources Association* 42 (1), 237–256.
- Xu, C.Y., Vandewiele, G.L., 1994. Sensitivity of monthly rainfall-runoff models to input errors and data length. *Hydrological Sciences Journal* 39 (2), 157–176.
- Yang, D., Goodison, B.E., Ishida, S., Benson, C.S., 1998a. Adjustment of daily precipitation data at 10 climate stations in Alaska: application of World Meteorological Organization intercomparison results. *Water Resources Research* 34 (2), 241–256.
- Yang, D. et al., 1998b. Accuracy of NWS 8" standard nonrecording precipitation gauge: results and application of WMO intercomparison. *Journal of Atmospheric and Oceanic Technology* 15 (1), 54–68.
- Yapo, P.O., Gupta, H.V., Sorooshian, S., 1996. Automatic calibration of conceptual rainfall-runoff models: sensitivity to calibration data. *Journal of Hydrology* 181 (1–4), 23–48.
- Young, C.B., Bradley, A.A., Krajewski, W.F., Kruger, A., Morrissey, M.L., 2000. Evaluating NEXRAD multisensor precipitation estimates for operational hydrologic forecasting. *Journal of Hydrometeorology* 1 (3), 241–254.
- Zhang, Y., Reed, S., Kitzmiller, D., 2010. Effects of retrospective gauge-based readjustment of multisensor precipitation estimates on hydrologic simulations. *Journal of Hydrometeorology* 12 (3), 429–443.